

裸子植物 DNA 条形码 ITS2 高通用性引物的筛选*

李 燕^{1,2}, 吴 丁³, 高连明^{1**}

(1 中国科学院昆明植物研究所生物多样性与生物地理学重点实验室, 云南 昆明 650201; 2 云南省农业科学院高山经济植物研究所, 云南 丽江 674100; 3 景德镇学院, 江西 景德镇 333000)

摘要: DNA 条形码技术是利用标准 DNA 片段进行准确快速鉴定物种的一种方法, 理想的 DNA 条形码片段应具有高通用性。虽然核糖体 DNA 内部转录间隔区 II (ITS2) 被建议作为种子植物有效的 DNA 条形码, 但目前裸子植物还没有通用性高的引物可用。为获得高通用性的 ITS2 引物, 本研究基于裸子植物 55 个属的 5.8S 基因的保守序列区设计了 3 个正向引物, 与已有的 ITS 反向引物组合, 组成了 7 对 ITS2 引物进行通用性的评价。选取了裸子植物 8 目、12 科和 40 属的 56 个种用于本文的研究。引物组合 5.8SR/ITS4、5.8SRa/ITS4 和 5.8SF2/S3R 因为在科水平评价中通用性低或者产生的 PCR 产物有双带, 因而排除在全部物种水平上进一步评价。其余 4 对引物 (GYM_5.8SF1/ITS4、GYM_5.8SF1/S3R、GYM_5.8SF2/ITS4 和 S2F/S3R) 在 56 个物种的 PCR 检测中, 均有 100% 的扩增率。基于 PCR 产物的亮度、序列质量和正反向引物覆盖率的综合评价, 建议引物 GYM_5.8SF2/ITS4 作为裸子植物条形码片段 ITS2 最好的通用引物。

关键词: DNA 条形码; 裸子植物; ITS2; 引物通用性; 物种鉴定

中图分类号: Q 781

文献标识码: A

文章编号: 2095-0845(2013)06-751-10

Highly Universal DNA Barcoding Primers of ITS2 for Gymnosperms*

LI Yan^{1,2}, WU Ding³, GAO Lian-Ming^{1**}

(1 Key Laboratory of Biodiversity and Biogeography, Kunming Institute of Botany, Chinese Academy of Sciences, Kunming 650201, China; 2 Institute of Alpine Economic Plants, Yunnan Academy of Agricultural Sciences, Lijiang 674100, China; 3 Jingdezhen University, Jingdezhen 333000, China)

Abstract: DNA barcoding is an approach for rapid and accurate species identification using a standard DNA region. Ideal barcodes should have high level of universality in PCR and sequencing. Although the nuclear ribosomal DNA (nrDNA) internal transcribed spacers 2 (ITS2) was proposed as potential DNA barcode for seed plants, including gymnosperms, no suitable ITS2 primer pair for gymnosperms available shows high universality in either aspect. To obtain a highly universal ITS2 primer pair, we newly designed three forward primers for ITS2 based on sequences in the 5.8S gene greatly conserved among 55 genera of gymnosperms. Combined with previously published reverse primers of ITS, seven candidate primer combinations for ITS2 were assessed for their universality. A total of 56 species, representing 40 genera from all the 12 families of the eight orders of gymnosperms, were sampled in this study. Three ITS2 primer combinations, 5.8SR/ITS4, 5.8SRa/ITS4, and 5.8SF2/S3R, were excluded due to their low universality or because of yielding double bands in the PCR in the family-level assessment. The remaining four primer combinations, GYM_5.8SF1/ITS4, GYM_5.8SF1/S3R, GYM_5.8SF2/ITS4, and S2F/S3R, showed high PCR universality (100%) on all 56 sampled species. Based on our overall assessment of PCR and sequencing universality

* Funding: The Ministry of Science and Technology of the People's Republic of China (No. 2012FY110800, 2013FY112600), the Large-scale Scientific Facilities of the Chinese Academy of Sciences (No. 2009-LSFGBOWS-01)

** Author for correspondence; E-mail: gaolm@mail.kib.ac.cn

Received date: 2013-09-20, Accepted date: 2013-09-30

作者简介: 李 燕 (1976-) 女, 副研究员, 主要从事高山经济植物的引种驯化和 DNA 条形码研究。E-mail: enleiyin@163.com

ty, brightness of PCR bands, and sequence quality and coverage, we propose the primer combination GYM_5.8SF2/ITS4 as ITS2 barcode for gymnosperms because it was shown to be the most “universal”.

Key words: DNA barcoding; Gymnosperms; ITS2; Universal primer; Species identification

DNA barcoding is an approach to identify species and/or discover new species using one or a few standardized DNA regions from different genomes (Hebert *et al.*, 2003; Hollingsworth, 2011; Gao *et al.*, 2012), which has been widely used in the field of taxonomy, ecological surveys and assessment of biodiversity (Hebert *et al.*, 2004; Liu *et al.*, 2011; Lahaye *et al.*, 2008; Valentini *et al.*, 2009; García-Robledo *et al.*, 2013). The ideal DNA barcodes should satisfy the criterion of an appropriately short sequence length (300–800 base pairs, bp) to facilitate PCR amplification with a single primer pair (universality), be amenable to PCR and bidirectional Sanger sequencing with little requirement for manual editing of the sequence trace files (sequence quality and coverage), and provide maximal discrimination among the species (Kress *et al.*, 2005; CBOL Plant Working Group, 2009). The mitochondrial gene *CYTOCHROME OXIDASE 1* (*COI*) has been proven to be a remarkably effective “animal barcode” for discriminating species of birds, fishes, and insects (Hebert *et al.*, 2004; Holmes *et al.*, 2009; Linares *et al.*, 2009; Hollingsworth, 2011). However, it is more difficult to find a similarly robust and effective barcode for plants (CBOL Plant Working Group, 2009; Hollingsworth *et al.*, 2011; Li *et al.*, 2011a). Although many studies have compared the performance of a range of candidate DNA regions as barcodes (e.g. Sass *et al.*, 2007; Fazekas *et al.*, 2008; Hollingsworth *et al.*, 2009; Seberg and Petersen, 2009; Li *et al.*, 2011b), no single candidate DNA locus evaluated fitted all criteria for seed plants. Thus, combinations of different DNA loci were suggested as barcodes for plants (Chase *et al.*, 2007; Kress and Erickson, 2007; Hollingsworth *et al.*, 2009; CBOL Plant Working Group, 2009).

Kress *et al.* (2005) initially recommended the cpDNA *trnH-psbA* intergenic spacer and the nuclear

ribosomal DNA (nrDNA) internal transcribed spacer (ITS) region as barcodes for flowering plants. In 2009, the Consortium for the Barcode of Life (CBOL) Plant Working Group proposed the combination of the coding regions of *rbcL*+*matK* as the core barcode for land plants. Thereafter, the combination of *trnH-psbA* and ITS2 (internal transcribed spacer 2) was demonstrated to represent an effective barcode for species identification of medicinal plants in several studies (e.g. Chen *et al.*, 2010; Pang *et al.*, 2010; Gao *et al.*, 2010). ITS2 was proposed to be used as a universal DNA barcode complementing *COI* to identify animal species (Yao *et al.*, 2010), and as complementary barcode to the core barcode of *rbcL*+*matK* for plants (Hollingsworth *et al.*, 2011). Subsequently, based on the assessment of a large scale dataset, the China Plant BOL Group (Li *et al.*, 2011b) backed the incorporation of ITS or ITS2 into the core barcode for seed plants.

Though many researchers have proposed the use of nrDNA ITS as a standard barcode for plants, there is a residual nervousness stemming from three major potential problems; i.e. fungal contamination, presence of paralogous copies, and recovery (simply difficult to amplify and sequence) (Hollingsworth, 2011). Of these, the main limitation is the sometimes great difficulty to amplify and sequence the ITS region (Kress *et al.*, 2005; Li *et al.*, 2011b), perhaps due to non-conserved primer sites (Möller, 2000). For instance, the recovery rate of 76.5% was some 10% – 15% lower than for the other barcode markers based on a 6286 sample dataset (Li *et al.*, 2011b). A high level of primer universality for polymerase chain reaction (PCR) and sequencing is one of the most important criteria for DNA barcoding (Chase *et al.*, 2007; Kress and Erickson, 2007; Ford *et al.*, 2009; Hollingsworth *et al.*, 2009). Thus, it is very important to select highly efficient primer pairs

for the PCR amplification of ITS or ITS2.

Gymnosperms are a class of seed-bearing plants, which are the dominant vegetation in many colder and arctic regions and have major economic uses as important wood resources and as ornamental plants. The class consists of four subclasses, i. e. Cycadidae, Ginkgoideae, Gnetidae, and Pinidae, representing eight orders, 12 families, 83 genera, and approximately 990 species (Christenhusz *et al.*, 2011). There are about 250 gymnosperm species in 34 genera and 10 families in China (Fu *et al.*, 1999). The great variation in length of the ITS region (including ITS1, ITS2 and 5.8S) in gymnosperms is possibly a limit for the use of the region as DNA barcode. However, length variation was predominantly observed for ITS1 (e. g. ranging from 630 to 3 125 bp in length in *Pinus*, Liston *et al.*, 1999), while the ITS2 and 5.8S gene are more length conserved (Liston *et al.*, 1996, 1999; Gernandt and Liston, 1999). Therefore, ITS2 seems a more likely candidate for a gymnosperm DNA barcode, when obtaining full ITS sequences is problematic. Another detrimental aspect of is the sometimes low universality of ITS when using universal gymnosperm PCR primers (Li *et al.*, 2011b), also reported for gymnosperm universal primers for ITS2 (Chiou *et al.*, 2007; Chen *et al.*, 2010). Thus, a highly PCR universal primer pair for ITS2 is required if this region is even to be considered suitable for gymnosperms.

In the present study, we therefor newly designed and tested three forward primers, GYM_5.8SF1, GYM_5.8SF2, and 5.8SRa, based on sequences in the 5.8S gene greatly conserved over 55 gymnosperm genera. We selected 56 species representing 42 genera of all 12 gymnosperm families to test the new primers. A total of seven candidate primer combinations for ITS2 were evaluated to determine their universality in PCR amplification across the 56 gymnosperm species and Sanger sequencing of seven species from the four subclasses of gymnosperms here. The ultimate aim was to be able to propose the most universal primer pairs of ITS2 for barcoding gymnosperms.

Materials and methods

Sample strategy

To represent the large diversity across gymnosperms, a total of 56 species representing 42 genera from 12 families of eight orders in gymnosperms were sampled in this study (Table 1), covering around 49% of genera, and all families, orders, and subclasses of gymnosperms. All gymnosperm species occurring in China were covered. Voucher specimens were deposited at the herbarium of Kunming Institute of Botany, Chinese Academy of Sciences (KUN). For some larger genera, more than one species per genus were selected.

DNA extraction

Total DNA was extracted from silica-gel dried leaf tissues using the CTAB method (Doyle and Doyle, 1987). The concentration of total genomic DNA was checked using spectrophotometer (ND-1000, NanoDrop, Wilmington, USA), and then was dissolved in TE buffer (10 mmol · L⁻¹ Tris-HCl, pH 8.0, 1 mmol · L⁻¹ EDTA) to a final concentration of 40–50 ng · μL⁻¹ to avoid any variation in PCR success due to DNA concentration differences.

Primer design and primer combination screening strategy

Sequences of the 5.8S gene and ITS2 for a total of 74 species, representing 55 genera from all the 12 families of gymnosperms, were obtained from GenBank. These sequences were aligned in Clustal X (Larkin *et al.*, 2007). Three new forward primers (GYM_5.8SF1, GYM_5.8SF2, and 5.8SRa, see Table 2) were designed in regions of the 5.8S gene largely conserved across the gymnosperms included here with the help of Oligo 6.0 (Offerman and Rychlik, 2003). A further forward primer, 5.8SR (designed by Vilgalys lab: <http://biology.duke.edu/fungi/mycolab/primers.htm>) was selected for the universality assessment in our study as well. In combination with the reverse primers ITS4 and S3R, we had six ITS2 primer combinations to test. The primer pair, S2F/S3R for ITS2 (Chiou *et al.*, 2007), proposed for plant barcoding in several studies previous-

ly (Chen *et al.*, 2010; Gao *et al.*, 2010; Pang *et al.*, 2010, 2012), was used as a bench mark in comparison to primer pairs including our newly designed primers (Table 2).

Table 1 Voucher information and sample code of the species included in this study, and polymerase chain reaction success for four ITS2 primer combinations

Subclass	Family	Taxon	Voucher	Code	Primer			
					GYM_5.8SF1 ITS4	GYM_5.8SF1 S3R	GYM_5.8SF2 ITS4	S2F S3R
Cycadidae	Cycadaceae	<i>Cycas guizhouensis</i> *	Glm-06032	G1	+++	+++	+++	+++
Cycadidae	Cycadaceae	<i>Cycas micholitzii</i>	Ly-012	G2	+++	+++	+++	+++
Cycadidae	Zamiaceae	<i>Encephalartos lehmannii</i> *	Ly-014	G3	+++	+++	+++	+++
Ginkgoidae	Ginkgoaceae	<i>Ginkgo biloba</i> *	Ly-028	G4	+	+	+++	++
Gnetidae	Ephedraceae	<i>Ephedra intermedia</i>	12384	G57	+++	+++	+++	+++
Gnetidae	Ephedraceae	<i>Ephedra likiangensis</i> *	Ly-024	G55	+++	+++	+++	+++
Gnetidae	Gnetaceae	<i>Gnetum pendulum</i> *	GBOWS0801	G54	+++	+++	+++	+++
Gnetidae	Welwitschiaceae	<i>Welwitschia mirabilis</i> *	RBGE19941311	G59	+++	+++	+++	+++
Pinidae	Araucariaceae	<i>Agathis dammara</i>	Gy058	G58	+++	+++	+++	+++
Pinidae	Araucariaceae	<i>Araucaria bidwillii</i>	Ly-015	G46	+++	+++	+++	+++
Pinidae	Araucariaceae	<i>Araucaria cunninghamii</i> *	Ly-016	G45	+++	+++	+++	+++
Pinidae	Araucariaceae	<i>Araucaria heterophylla</i>	Ly-004	G47	+++	+++	+++	+++
Pinidae	Cupressaceae	<i>Calocedrus macrolepis</i> var. <i>formosana</i>	Glm-103074	G35	+++	+++	+++	+++
Pinidae	Cupressaceae	<i>Chamaecyparis formosensis</i>	Glm-103065	G33	+++	+++	+++	+++
Pinidae	Cupressaceae	<i>Cunninghamia lanceolata</i>	Ly-007	G22	++	++	++	++
Pinidae	Cupressaceae	<i>Cupressus duclouxiana</i>	Ly-011	G32	+++	+++	+++	+++
Pinidae	Cupressaceae	<i>Cryptomeria japonica</i>	Glm-103142	G25	+++	+++	+++	+++
Pinidae	Cupressaceae	<i>Fokienia hodginsii</i>	Ly-006	G31	+++	+++	+++	+++
Pinidae	Cupressaceae	<i>Glyptostrobus pensilis</i> *	Ly-005	G26	+++	+++	+++	+++
Pinidae	Cupressaceae	<i>Juniperus pingii</i> var. <i>wilsonii</i>	Ly-026	G38	+++	+++	+++	+++
Pinidae	Cupressaceae	<i>Juniperus squamata</i> *	Glm-082126	G30	+++	+++	+++	+++
Pinidae	Cupressaceae	<i>Metasequoia glyptostroboides</i>	Ly-002	G29	+++	+++	+++	+++
Pinidae	Cupressaceae	<i>Platycladus orientalis</i>	Ly-020	G37	+++	+++	+++	+++
Pinidae	Cupressaceae	<i>Sequoia sempervirens</i>	Ly-021	G28	+++	+++	+++	+++
Pinidae	Cupressaceae	<i>Taiwania flousiana</i>	Glm-092411	G24	+++	+++	+++	+++
Pinidae	Cupressaceae	<i>Taxodium distichum</i>	Ly-009	G27	+++	+++	+++	+++
Pinidae	Cupressaceae	<i>Thujopsis dolabrata</i>	Ly-019	G34	+++	+++	+++	+++
Pinidae	Cupressaceae	<i>Thuja sutchuenensis</i>	Ly-018	G36	+++	+++	+++	+++
Pinidae	Pinaceae	<i>Abies georgei</i>	Glm-102904	G12	+++	+++	+++	+++
Pinidae	Pinaceae	<i>Cathaya argyrophylla</i> *	Ly-017	G13	+++	+++	+++	+++
Pinidae	Pinaceae	<i>Cedrus deodara</i>	Ly-029	G8	+++	+++	+++	+++
Pinidae	Pinaceae	<i>Keteleeria davidiana</i>	Ly-001	G14	+++	+++	+++	+++
Pinidae	Pinaceae	<i>Keteleeria evelyniana</i>	Ly-003	G15	+++	+++	+++	+++
Pinidae	Pinaceae	<i>Keteleeria fortunei</i>	Ly-010	G16	+++	+++	+++	+++
Pinidae	Pinaceae	<i>Larix himalaica</i>	Glm-081604	G10	+++	+++	+++	+++
Pinidae	Pinaceae	<i>Larix potaninii</i> var. <i>chinensis</i>	Glm-06081	G9	+++	+++	+++	+++
Pinidae	Pinaceae	<i>Picea likiangensis</i>	Glm-08912	G19	+++	+++	+++	+++
Pinidae	Pinaceae	<i>Picea smithiana</i>	Glm-081533	G18	+++	+++	+++	+++
Pinidae	Pinaceae	<i>Pinus armandi</i>	Ly-027	G7	+++	+++	+++	+++
Pinidae	Pinaceae	<i>Pinus wallichiana</i>	Glm-081471	G5	+++	+++	+++	+++
Pinidae	Pinaceae	<i>Pinus yunnanensis</i> *	Ly-025	G6	+++	+++	+++	+++
Pinidae	Pinaceae	<i>Pseudolarix amabilis</i>	Ly-008	G11	+++	+++	+++	+++

Table 1 continued

Subclass	Family	Taxon	Voucher	Code	Primer			
					GYM_ 5.8SF1 ITS4	GYM_ 5.8SF1 S3R	GYM_ 5.8SF2 ITS4	S2F S3R
Pinidae	Pinaceae	<i>Pseudotsuga sinensis</i>	Glm-103063	G17	+++	+++	+++	+++
Pinidae	Pinaceae	<i>Tsuga dumosa</i>	Glm-081797	G21	+++	+++	+++	+++
Pinidae	Podocarpaceae	<i>Dacrycarpus imbricatus</i>	Ly-023	G39	+++	+++	+++	+++
Pinidae	Podocarpaceae	<i>Nageia fleuryi</i> *	Ly-022	G40	+++	+++	+++	+++
Pinidae	Podocarpaceae	<i>Nageia nagi</i>	Ly-013	G41	+++	+++	+++	+++
Pinidae	Podocarpaceae	<i>Podocarpus macrophyllum</i>	GBOWS0294	G42	+++	+++	+++	+++
Pinidae	Podocarpaceae	<i>Podocarpus neriifolius</i>	Glm06211	G43	+++	+++	+++	+++
Pinidae	Sciadopityaceae	<i>Sciadopitys verticillata</i> *	Glm-092241	G44	+++	+++	+++	+++
Pinidae	Taxaceae	<i>Amentotaxus argotaenia</i>	182	G48	+++	+++	+++	+++
Pinidae	Taxaceae	<i>Cephalotaxus mannii</i>	ZSD001	G53	+++	+++	+++	+++
Pinidae	Taxaceae	<i>Cephalotaxus wilsoniana</i>	Glm-103119	G52	+++	+++	+++	+++
Pinidae	Taxaceae	<i>Pseudotaxus chienii</i> *	Glm-07549	G51	+++	+++	+++	+++
Pinidae	Taxaceae	<i>Taxus wallichiana</i>	RC1289	G50	+++	+++	+++	+++
Pinidae	Taxaceae	<i>Torreya fargesii</i> var. <i>yunnanensis</i>	Glm-092567	G49	+++	+++	+++	+++

Species in bold were selected for sequencing to evaluate sequence quality and coverage in this study. * Species were selected for evaluating PCR success using the seven candidate ITS2 primer combinations at family-level. +++, strong band; ++, medium band; +, weak band

Table 2 The seven combinations of ITS2 primers tested for universality in this study

Primer name	Direction	Sequence (5'-3')	Length /bp	Optimum annealing temperature/°C	Source
GYM_5.8SF1 ITS4	F R	TTGYAGAATCCCGTGARTC TCCTCCGCTTATTGATATGC	19 20	48	Designed in this study White <i>et al.</i> , 1990
GYM_5.8SF1 S3R	F R	TTGYAGAATCCCGTGARTC GACGCTTCTCCAGACTACAAT	19 21	50	Designed in this study Chiou <i>et al.</i> , 2007
GYM_5.8SF2 ITS4	F R	GYAGAATCCCGTGARTCATC TCCTCCGCTTATTGATATGC	20 20	52	Designed in this study White <i>et al.</i> , 1990
GYM_5.8SF2 S3R	F R	GYAGAATCCCGTGARTCATC GACGCTTCTCCAGACTACAAT	20 21	52	Designed in this study Chiou <i>et al.</i> , 2007
5.8SR ITS4	F R	TCGATGAAGAACGCAGCG TCCTCCGCTTATTGATATGC	18 20	52	Vilgalys lab White <i>et al.</i> , 1990
5.8SRa ITS4	F R	WCGATGAAGAACGYAGCG TCCTCCGCTTATTGATATGC	18 20	52	Designed in this study White <i>et al.</i> , 1990
S2F S3R	F R	ATGCGATACTTGCTGTGAAT GACGCTTCTCCAGACTACAAT	20 21	52	Chiou <i>et al.</i> , 2007 Chiou <i>et al.</i> , 2007

F, forward; R, reverse. R-A/G, W-A/T, Y-C/T; bp-base combinations

To determine the optimum annealing temperature, gradient PCR (annealing temperature ranging from 46 °C to 58 °C with a 2 °C gradient) was performed using the seven primer combinations on four species, *Cycas guizhouensis*, *Ginkgo biloba*, *Pinus yunnanensis*, and *Gnetum pendulum*, representing the four gymnosperm subclasses (Table 1). The gradient PCR reactions were carried out twice to con-

firm the optimum annealing temperatures. We then adopted a two-step approach to determine the universality of the six primer combinations (Li *et al.*, 2011). First, we selected 14 species representing all the 12 families for PCR amplification with all the candidate ITS2 primer combinations. For the large families Pinaceae and Cupressaceae, two species were used in the family-level assessment (Table 1).

Second, the candidate primer combinations performed well with a single PCR band and high PCR success rate were selected for further assessment on all the remaining sampled species.

PCR amplification and sequencing

All PCR reactions were carried out with a 2720 Thermal Cycler (Applied Biosystems, Foster City, CA, USA). The final reaction volume of 20 μL contained 0.5 μL each of the forward and reverse primers (5 $\mu\text{mol} \cdot \text{L}^{-1}$), 10 μL 2 $\times$ Taq PCR Master Mix (0.1 U Taq polymerase $\cdot \mu\text{L}^{-1}$, 0.5 $\text{mmol} \cdot \text{L}^{-1}$ each dNTP, 20 $\text{mmol} \cdot \text{L}^{-1}$ Tris-HCl (pH 8.3), 100 $\text{mmol} \cdot \text{L}^{-1}$ KCl, 3 $\text{mmol} \cdot \text{L}^{-1}$ MgCl_2 ; Tiangen Biotech, Beijing, China), 0.2 μL bovine serum albumin (10 $\mu\text{g} \cdot \mu\text{L}^{-1}$) and 1 μL template DNA. Negative controls were run alongside all PCR reactions. The thermocycling profile for ITS2 was as follows: 94 $^{\circ}\text{C}$ for 3 min, then 35 cycles of 94 $^{\circ}\text{C}$ for 1 min, 46–58 $^{\circ}\text{C}$ (in the gradient PCR) or the primer pair specific optimum annealing temperature (Table 2) for 45 s, 72 $^{\circ}\text{C}$ for 1 min, and a final extension step of 72 $^{\circ}\text{C}$ for 7 min. The PCR products were analyzed by electrophoresis on 1.2% Tris-acetate-EDTA agarose gels stained with ethidium bromide alongside a GeneRuler 100 bp DNA ladder (Fermentas, Glen Burnie, MD, USA).

The PCR products were purified using ExoSAP-IT (GE Healthcare, Cleveland, OH, USA). Sequencing reactions were carried out on forward and reverse strands in a total volume of 6 μL containing 0.2 μL of the purified PCR product, 0.15 μL of BigDye terminator sequencing mixture (V3.1), 1.2 μL of sequence buffer, and 1.4 μM of primer). The Sanger sequencing profile was 32 cycles of 96 $^{\circ}\text{C}$ for 10 s, 50 $^{\circ}\text{C}$ for 5 s, and 60 $^{\circ}\text{C}$ for 4 min. The purified sequencing products were run on an ABI 3730xl automated sequencer (Applied Biosystems, Foster City, USA).

Analysis of PCR and sequencing success, sequence quality and coverage

Recording of the PCR and sequencing success followed the methods of Li *et al.* (2011b). The se-

quence quality was assessed using the software Sequencing Analysis 5.3.1 (Applied Biosystems, Foster City, USA) and followed Li *et al.* (2011). The percentage of bases with >QV30 (QV, an established metric for determining quality sequencing data) and the coverage of the bidirectional sequences was determined using Sequencher 5.0 (Gene Codes Corporation, Ann Arbor, MI, USA).

Results

Optimum annealing temperature of the seven candidate primer combinations

The optimum annealing temperature for the candidate primer combination was determined based on the success rate of the PCR reactions and the brightness of the PCR bands (assumed to reflect amplification efficiency). In this study, the optimum annealing temperature for the primer combinations GYM_5.8SF1/ITS4 and GYM_5.8SF1/S3R was 48 $^{\circ}\text{C}$ and 50 $^{\circ}\text{C}$, respectively. The remaining five primer combinations showed the same optimum annealing temperature of 52 $^{\circ}\text{C}$ (Table 1).

Universality of the candidate primer combinations

In the universality assessment at the family-level on 14 samples, primer pair 5.8SR/ITS4 yielded strong PCR bands for seven samples (50%), weak PCR bands for five samples (35.7%), and no bands for two species (*Ginkgo biloba* and *Sciadopitys verticillata*). Though the primer combinations of 5.8SRa/ITS4 and GYM_5.8SF2/S3R produced bands for all 14 samples (100%), the PCR products for some species showed double bands on the agarose gel. Because of the low universality and the amplification of more than one band, these three primer pair were excluded from further assessments. The remaining four ITS2 primer combinations (GYM_5.8SF1/ITS4, GYM_5.8SF1/S3R, GYM_5.8SF2/ITS4, and S2F/S3R) showed 100% PCR success (Table 1), though the primer combinations GYM_5.8SF1/ITS4 and GYM_5.8SF1/S3R produced a weak PCR band for *Ginkgo biloba* (Table 1). The primer combination GYM_5.8SF2/ITS4 yielded a strong PCR

band and S2F/S3R a medium bright PCR band for *Ginkgo biloba* (G4), respectively (Figure 1). The remaining sampled species showed a single strong PCR band on the family-level assessment for the four primer combinations (Figure 1, Table 1).

The universality of the four ITS2 candidate primer combinations performed uniformly well on the remaining 42 sampled species with 100% PCR success and a

single strong PCR band per reaction except for *Cunninghamia lanceolata* (G22) for which consistently medium bright PCR band were obtained (Figure 1, Table 1). Thus, the overall universality assessment across all 56 species for the four candidate primer combinations was 100% PCR success. Considering the band intensities, primer combination GYM_5.8SF2/ITS4 performed best, followed by the pair S2F/S3R.

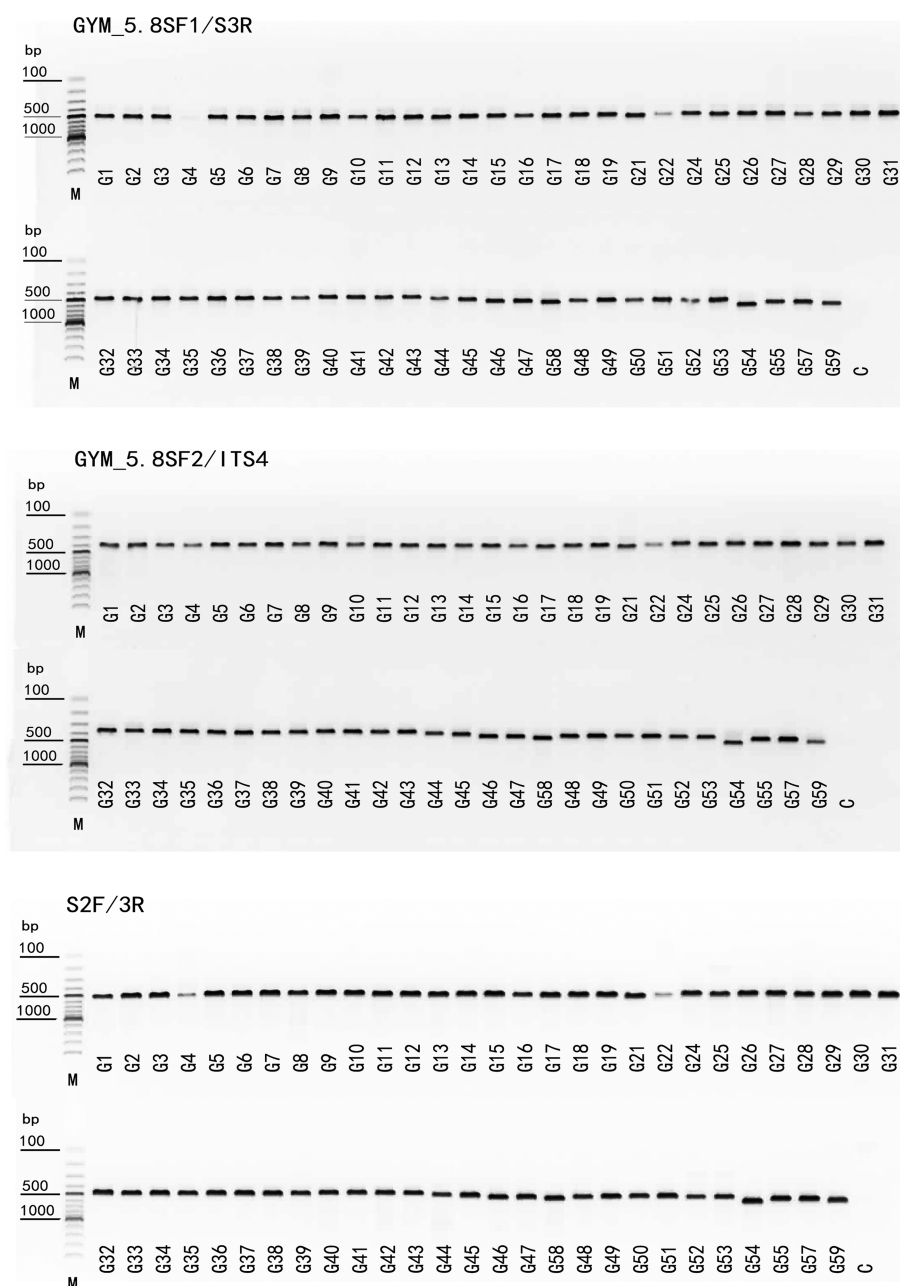


Fig. 1 PCR amplification profiles of ITS2 with primer combinations GYM_5.8SF1/S3R, GYM_5.8SF2/ITS4, and S2F/S3R, on all 56 species sampled for this study. Sample code as in Table 1. C means negative control for PCR, M means DNA ladder (100 bp)

Sequencing success and sequence quality

The assessment of the sequencing universality and the quality of the generated sequence traces for the two best PCR performing ITS2 primer combinations, GYM_5.8SF2/ITS4 and S2F/S3R, demonstrated 100% sequencing success with the seven species tested. All the 28 generated sequence traces of the seven tested species were of high base call quality, with few low-quality bases needing trimming out. The complete amplicon length for the two primer combinations ranged from 380 bp (GYM_5.8SF2/ITS4) to 510 bp (S2F/S3R), including primer sites, for *Cycas guizhouensis*. When using the default settings for sequence trimming in Sequencher 91.2–99.4% and 95.5–99.5% of the bases (QV>30) remained with primer combination GYM_5.8SF2/ITS4 and S2F/S3R, respectively. Thus, the sequence quality was very high and similar for the two primer combinations. The assembled contigs showed >70% overlap in the alignment of the forward and reverse reads without low-quality base calls for any of the generated ITS2 sequences.

Discussion

The nuclear ribosomal DNA (nrDNA) internal transcribed spacer (ITS) region is one of the most used DNA region for molecular systematics and evolution (Álvarez and Wendel, 2003; Kress *et al.*, 2005; Nieto *et al.*, 2008). It can provide significant sequence variability at the species level or lower (Yao *et al.*, 2010; Liu *et al.*, 2011). Because of the greater discriminatory power of nrITS over plastid regions at low taxonomic levels among plant species, it has been proposed as a DNA barcode in several studies (Roy *et al.*, 2010; Li *et al.*, 2011b; Muellner *et al.*, 2011; Simon *et al.*, 2012). The remarkable length variation of ITS in gymnosperms (Liston *et al.*, 1996, 1999; Gernandt and Liston 1999) is an obvious obstacle for a DNA barcode (Liu *et al.*, 2011). The “universal” ITS primer pair for gymnosperms has revealed low level of PCR and sequencing success (Li *et al.*, 2011b). It limited the use

of ITS as a barcode in gymnosperms even though it was proposed as a core barcode for seed plants (Li *et al.*, 2011b). However, ITS2, the second spacer of the ITS region, has been suggested as an effective DNA barcode for plants (Chen *et al.*, 2010; Pang *et al.*, 2010; Yao *et al.*, 2010; García-Robledo *et al.*, 2013). If ITS in its entirety is difficult to amplify and directly sequence, ITS2 could be an alternative for gymnosperms because it is shorter and easier to sequence than the entire ITS region (Coleman *et al.*, 2003). Therefore, ITS2 can represent a useful back-up barcode of ITS (Li *et al.*, 2011b).

The ITS2 primer pair used for gymnosperms showed relative low PCR and/or sequencing success in previous study (Chen *et al.*, 2010; Pang *et al.*, 2012). It is therefore the lack of a highly universal primer pair for ITS2 that prevents its wider use in gymnosperms. Thus, primer universality is one of the first to fulfilling criteria when screening for a DNA barcode (Chase *et al.*, 2007; Kress and Erickson, 2007; Ford *et al.*, 2009; Hollingsworth *et al.*, 2009). In this study, we designed three new primers based on sequences of the 5.8S gene greatly conserved across 55 genera of gymnosperms. When the polymorphic site exists within the primers region, we used the combined base in the prime (Table 2). Four of the seven candidate ITS2 primer combinations here performed very well with 100% PCR success for all the tested 56 species covering all the families of the eight orders of gymnosperms. Only *Ginkgo biloba* and *Cunninghamia lanceolata* yielded weak or medium strong PCR bands for the four primer combinations. *Ginkgo biloba* is the only living representative of the family Ginkgoaceae with an isolated systematic position in the gymnosperms. This distant genetic relationship of *Ginkgo biloba* with other groups of gymnosperms may have resulted in changes in the annealing sites for most combinations (except for GYM_5.8SF2/ITS4) and may explain the weak or medium PCR band of the candidate ITS2 primers for this species (*c.f.* Möller, 2000). The case for *Cunninghamia lanceolata* (Cupressaceae)

is likely different, because all other sampled species of Cupressaceae yielded strong PCR bands. A relatively poor quality of template DNA used here or the presence of inhibitors might be responsible for the consistently weak PCR bands of *Cunninghamia lanceolata*.

Based on the PCR success and yield of the PCR products of the four ITS2 primer combinations fully tested, two combinations GYM_5.8SF2/ITS4 and S2F/S3R performed slightly better. Of these two GYM_5.8SF2/ITS4 performed slightly better than the benchmark primer pair S2F/S3R, and yielded strong PCR band for most sampled species here. Both primer combinations, GYM_5.8SF2/ITS4 and S2F/S3R, performed equally well in their universality of sequencing. The generated sequences showed high levels of sequence quality and coverage as well, which is also an important criterion for DNA barcode screening. Based on our overall assessments of primer universality and sequence quality and coverage, we propose GYM_5.8SF2/ITS4 as the most “universal” primer pair for ITS2 for barcoding in gymnosperms.

Acknowledgements: We are grateful to QL Zhang, JB Yang, J Yang, HT Li and LJ Yan for their help in lab work. We also thank Michael Möller from Royal Botanic Garden Edinburgh for improving English of the manuscript. Laboratory work was performed at the Laboratory of Molecular Biology at the Germplasm Bank of Wild Species, Kunming Institute of Botany, Chinese Academy of Sciences.

References:

- Álvarez I, Wendel JF, 2003. Ribosomal ITS sequences and plant phylogenetic inference [J]. *Molecular Phylogenetics and Evolution*, **29**: 417—434
- CBOL Plant Working Group, 2009. A DNA barcode for land plants [J]. *Proceedings of the National Academy of Sciences of the United States of America*, **106**: 12794—12797
- Chase MW, Cowan RS, Hollingsworth PM *et al.*, 2007. A proposal for a standardised protocol to barcode all land plants [J]. *Taxon*, **56**: 295—299
- Chen SL, Yao H, Han JP *et al.*, 2010. Validation of the ITS2 region as a novel DNA barcode for identifying medicinal plant species [J]. *PLoS ONE*, **5**: e8613.
- Chiou SJ, Yen JH, Fang CL *et al.*, 2007. Authentication of medicinal herbs using PCR-amplified ITS2 with specific primers [J]. *Planta Medica*, **73**: 1421—1426
- Christenhusz MJM, Gardner MF, Mill RR *et al.*, 2011. A new classification and linear sequence of extant gymnosperms [J]. *Phytotaxa*, **19**: 55—70
- Coleman AW, 2003. ITS2 is a double-edged tool for eukaryote evolutionary comparisons [J]. *Trends in Genetics*, **19**: 370—375
- Doyle JJ, Doyle JL, 1987. A rapid DNA isolation procedure for small quantities of fresh leaf tissue [J]. *Phytochemical Bulletin*, **19**: 11—15
- Fazekas AJ, Burgess KS, Kesanakurti PR *et al.*, 2008. Multiple multilocus DNA barcodes from the plastid genome discriminate plant species equally well [J]. *PLoS ONE*, **3**: e2802
- Ford CS, Ayres KL, Toomey N *et al.*, 2009. Selection of candidate coding DNA barcoding regions for use on land plants [J]. *Botanical Journal of the Linnean Society*, **159**: 1—11
- Fu LG, Li N, Mill R, 1999. Taxaceae [A]. In: Wu ZY, Raven PH (eds.), *Flora of China* [M]. Beijing: Science Press; St. Louis, Missouri: Missouri Botanical Garden, **4**: 89—96
- Gao LM (高连明), Liu J (刘杰), Cai J (蔡杰) *et al.*, 2012. A synopsis of technical notes on the standards for plant DNA Barcoding [J]. *Plant Diversity and Resources* (植物分类与资源学报), **34**: 592—606
- Gao T, Yao H, Song J *et al.*, 2010. Evaluating the feasibility of using candidate DNA barcodes in discriminating species of the large Asteraceae family [J]. *BMC Evolutionary Biology*, **10**: 324
- García-Robledo C, Erickson DL, Staines CL *et al.*, 2013. Tropical plant-herbivore networks: reconstructing species interactions using DNA barcodes [J]. *PLoS ONE*, **8**: e52967
- Gernandt DS, Liston A, 1999. Internal transcribed spacer region evolution in *Larix* and *Pseudotsuga* (Pinaceae) [J]. *American Journal of Botany*, **86**: 711—723
- Hebert PDN, Cywinska A, Ball SL, 2003. Biological identifications through DNA barcodes [J]. *Proceedings of the Royal Society of London. Series B; Biological Sciences*, **270**: 313—321
- Hebert PDN, Penton EH, Burns JM *et al.*, 2004. Ten species in one; DNA barcoding reveals cryptic species in the neotropical skipper butterfly *Astraptes fulgerator* [J]. *Proceedings of the National Academy of Sciences of the United States of America*, **101**: 14812—14817
- Hollingsworth ML, Clark A, Forrest LL *et al.*, 2009. Selecting barcoding loci for plants: evaluation of seven candidate loci with species-level sampling in three divergent groups of land plants [J]. *Molecular Ecology Resources*, **9**: 439—457
- Hollingsworth PM, Graham SW, Little DP, 2011. Choosing and using a plant DNA barcode [J]. *PLoS ONE*, **6**: e19254
- Hollingsworth PM, 2011. Refining the DNA barcode for land plants [J]. *Proceedings of the National Academy of Sciences of the United States of America*, **108**: 10371—10376

- ed States of America*, **108**: 19451—19452
- Holmes FA, Walters RS, Theriault RL *et al.*, 1991. Phase II trial of taxol, an active drug in the treatment of metastatic breast cancer [J]. *Journal of the National Cancer Institute*, **83**: 1797—1805
- Kress WJ, Wurdack KJ, Zimmer EA *et al.*, 2005. Use of DNA barcodes to identify flowering plants [J]. *Proceedings of the National Academy of Sciences of the United States of America*, **102**: 8369—8374
- Kress WJ, Erickson DL, 2007. A two-locus global DNA barcode for land plants: the coding *rbcL* gene complements the non-coding *trnH-psbA* spacer region [J]. *PLoS ONE*, **2**: e508
- Lahaye R, Van Der Bank M, Bogarin D *et al.*, 2008. DNA barcoding the floras of biodiversity hotspots [J]. *Proceedings of the National Academy of Sciences of the United States of America*, **105**: 2923—2928
- Larkin MA, Blackshields G, Brown NP *et al.*, 2007. Clustal W and clustal X version 2.0 [J]. *Bioinformatics*, **23**: 2947—2948
- Li DZ, Liu JQ, Chen ZD *et al.*, 2011a. Plant DNA barcoding in China [J]. *Journal of Systematics and Evolution*, **49**: 165—168
- Li DZ, Gao LM, Li HT *et al.*, 2011b. Comparative analysis of a large dataset indicates that internal transcribed spacer (ITS) should be incorporated into the core barcode for seed plants [J]. *Proceedings of the National Academy of Sciences of the United States of America*, **108**: 19641—19646
- Li Y, Gao LM, Poudel RC *et al.*, 2011. High universality of *matK* primers for barcoding gymnosperms [J]. *Journal of Systematics and Evolution*, **49**: 169—175
- Linares MC, Soto-Calderon ID, Lees DC *et al.*, 2009. High mitochondrial diversity in geographically widespread butterflies of Madagascar: A test of the DNA barcoding approach [J]. *Molecular Phylogenetics and Evolution*, **50**: 485—495
- Liston A, Robinson WA, Oliphant JM *et al.*, 1996. Length variation in the nuclear ribosomal DNA internal transcribed spacer region of non-flowering seed plants [J]. *Systematic Botany*, **21**: 109—120
- Liston A, Robinson WA, Pinero D *et al.*, 1999. Phylogenetics of *Pinus* (Pinaceae) based on nuclear ribosomal DNA internal transcribed spacer region sequences [J]. *Molecular Phylogenetics and Evolution*, **11**: 95—109
- Liu J, Möller M, Gao LM *et al.*, 2011. DNA barcoding for the discrimination of Eurasian yews (*Taxus* L., Taxaceae) and the discovery of cryptic species [J]. *Molecular Ecology Resources*, **11**: 89—100
- Möller M, 2000. How universal are universal rDNA primers? A cautionary note for plant systematists and phylogeneticists [J]. *Edinburgh Journal of Botany*, **57** (2): 151—156
- Muellner AN, Schaefer H, Lahaye R, 2011. Evaluation of candidate DNA barcoding loci for economically important timber species of the mahogany family (Meliaceae) [J]. *Molecular Ecology Resources*, **11**: 450—460
- Nieto Feliner G, Rosselló JA, 2007. Better the devil you know? Guidelines for insightful utilization of nrDNA ITS in species-level evolutionary studies in plants [J]. *Molecular Phylogenetics and Evolution*, **44**: 911—919
- Offerman J, Rychlik W, 2003. Oligo primer analysis software [A]. In: Krawetz S, Womble D (eds.), *Introduction to Bioinformatics: A Theoretical and Practical Approach* [M]. Totowa: Humana Press, New Jersey, USA, 345—361
- Pang XH, Luo H, Sun C *et al.*, 2012. Assessing the potential of candidate DNA barcodes for identifying non-flowering seed plants [J]. *Plant Biology*, **14**: 839—844
- Pang XH, Song JY, Zhu YJ *et al.*, 2010. Applying plant DNA barcodes for Rosaceae species identification [J]. *Cladistics*, **27**: 165—170
- Roy S, Tyagi A, Shukla V *et al.*, 2010. Universal plant DNA barcode loci may not work in complex groups: A case study with Indian *Berberis* species [J]. *PLoS ONE*, **5**: e13674
- Sass C, Little DP, Stevenson DW *et al.*, 2011. DNA barcoding in the cycadales: testing the potential of proposed barcoding markers for species identification of cycads [J]. *PLoS ONE*, **2**: e1154
- Seberg O, Petersen G, 2009. How many loci does it take to DNA barcode a crocus? [J]. *PLoS ONE*, **4** (2): e4598
- Simon UK, Trajanoski S, Kroneis T *et al.*, 2012. Accession-specific haplotypes of the internal transcribed spacer region in *Arabidopsis thaliana*—a means for barcoding populations [J]. *Molecular Biology and Evolution*, **29**: 2231—2239
- Valentini A, Pompanon F, Taberlet P, 2009. DNA barcoding for ecologists [J]. *Trends in Ecology and Evolution*, **24**: 110—117
- White TJ, Bruns T, Lee S *et al.*, 1990. Amplification and direct sequencing of fungal ribosomal RNA genes for phylogenetics [A]. In: Innis MA, Gelfand DH, Sninsky JJ *et al.* (eds.), *PCR Protocols: A Guide to Methods and Applications* [M]. New York: Academic Press, Inc
- Yao H, Song JY, Liu C *et al.*, 2010. Use of ITS2 region as the universal DNA barcode for plants and animals [J]. *PLoS ONE*, **5**: 370—375